

Shape Expression Schema の下での  
パターン問合せ充足可能性判定アルゴリズム  
An Algorithm for Detecting Pattern Queries  
under Shape Expression Schema

学籍番号：201921645

氏名：松岡 栞

Matsuoka Shiori

近年、RDF/グラフデータ(以下、グラフデータ)が急速に増加している。グラフデータの普及に伴い、既存のスキーマではスキーマ言語としての表現力が不十分であったため、グラフデータの構造を記述するための新しいスキーマ言語が必要とされた。そこで Shape Expression Schema (ShEx) が W3C Draft Community Group で検討されている。ShEx はグラフデータの構造的特徴を捉えられるように設計されており、すでにさまざまな分野で使用されている。グラフデータは規模が大きいくことが多く、より効率の良い問合せ処理が必要とされている。パターン問合せ  $q$  と ShEx スキーマ  $S$  が存在するときに、 $S$  に妥当かつ  $q$  の問合せ結果が空の場合、 $q$  は  $S$  の下で充足不能であると言う。充足不能な問合せの結果は空であり、その問合せ時間は無駄なものになってしまう。DTD や XML スキーマでの XPath 充足可能性問題に関する多くの研究が行われている。しかし、著者の知る限り、ShEx スキーマでのパターン問合せの充足可能性に関する研究はこれまで行われていない。

パターンが充足可能か否かを判定する方法として、部分グラフ同型問題を用いてパターンが ShEx スキーマにマッチするか否かを調べるという方法も考えられる。しかし、ShEx は選言「|」を用いて「 $E_1 | E_2$ 」のように  $E_1$  または  $E_2$  に一致するという型を定義できる等の特徴があるため、部分グラフ同型問題を解くだけでは不十分である。例えば、部分グラフ同型問題をそのまま適用した場合、分岐しているうちの一方のエッジが選択された時点で、他方のエッジは以降のマッチングにおいて利用不可能となることを十分に考慮できない。そこで本研究では、グラフデータと比較して容量の著しく軽い ShEx を参照し、パターンと同型のグラフを生成することでパターン問合せの充足可能性を判定するアルゴリズムを提案する。より具体的には部分グラフ同型問題の解法の一つである Ullman Algorithm を基に候補ノードを求め、その後マッチングを進める。この際、選言「|」による分岐をもつノードである場合は同時に分岐するエッジを使用するか否か、回数制限のあるエッジを使用する場合は制限にかからないか否か等を判定する。最終的にパターンが生成可能であれば充足可能、不可能であれば充足不能と出力するアルゴリズムである。本研究の目的が達成されれば大規模なグラフデータをより効率よく問合せることが可能となり、グラフデータを利用する研究者・開発者にとって有用と考えられる。

評価実験として、データセットと充足不能なクエリを用意し、各データセットに対してパターン問合せ、各データセットの ShEx スキーマに対して提案手法を実行した。評価実験の結果、提案アルゴリズムが充足可能性判定を行えること、充足不能場合に処理時間を少なく済ませられることを確認した。また、処理時間は実際のクエリ実行時間よりもはるかに短い時間で済むことが確認された。したがって、充足可能なクエリであっても充足可能性判定を事前に行うことで大きな時間の損失が発生することはないといえる。

研究指導教員：鈴木 伸崇  
副研究指導教員：永森 光晴